Supplementary Material B

Structural forgetting predictions for deeper embeddings

In Section 4.5, we showed conditions under which lossy-context surprisal with deletion noise predicts structural forgetting in a toy grammar. We presented predictions at embedding depths 1 and 2, indicating the presence of 1 or 2 nested relative clauses in the sentence prefix. It would also be possible to consider people's behavior given sentences with embedding depth 3 or higher:

(6) a. Embedding depth 1: The apartment$_1$ that the maid$_2$ cleaned$_2$ was well-decorated$_1$.

b. Embedding depth 2: The apartment$_1$ that the maid$_2$ who the cleaning service$_3$ sent over$_3$ cleaned$_2$ was well-decorated$_1$.

c. Embedding depth 3: The apartment$_1$ that the maid$_2$ who the cleaning service$_3$ that the manager$_4$ worked for$_4$ sent over$_3$ cleaned$_2$ was well-decorated$_1$.

Here we show that lossy-context surprisal predicts that structural forgetting will occur for these sentences even in languages with syntactic properties like German (i.e. consistently verb-final relative clauses). Figure B1 shows regions of different model behavior with respect to grammar and noise parameters, as in Figure 4, where now we have allowed the grammar to generate self-embeddings of depth 3.

In Figure B1, the pink region corresponds to grammars and noise models where there will be a structural forgetting effect at embedding depth 3 but not at embedding depth 1 or 2. German grammar is modelled by $f = 1$, the far right edge of each panel, which has a sizeable pink region when the deletion rate d is high and when the relative clause rate r is low—though note for $d = .1$ we would not get structural forgetting even for three levels of embedding if $f = 1$. If the overall relative clause rate in German is sufficiently low, then we broadly predict structural forgetting at embedding depth 3 in German, though it may be rare (as it requires a deletion rate of at least .2).

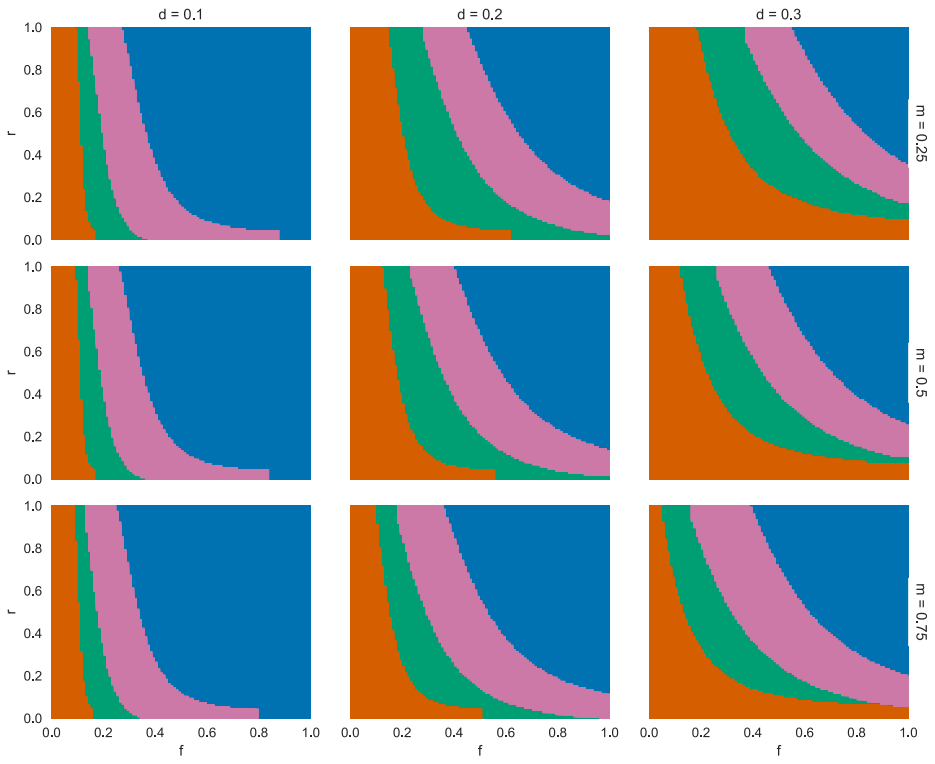

Figure B1. Regions of different model behavior regarding structural forgetting in terms of the free parameters d (noise rate), m (postnominal modification rate), r (relative clause rate), and f (verb-final relative clause rate), *for three levels of embedding*. = $\mathrm{U}_1\mathrm{U}_2\mathrm{U}_3$, = $\mathrm{G}_1\mathrm{U}_2\mathrm{U}_3$, = $\mathrm{G}_1\mathrm{G}_2\mathrm{U}_3$, = $\mathrm{G}_1\mathrm{G}_2\mathrm{G}_3$.